MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 6: 59–74
DOI 10.3897/mbmg.6.79265

Research Article

Metagenomic insights to the functional potential of sediment microbial communities in freshwater lakes

Laura Biessy[1*], **John K. Pearman**[1*], **Sean Waters**[1], **Marcus J. Vandergoes**[2], **Susanna A. Wood**[1]

1 *Cawthron Institute, Coastal and Freshwater Group, 98 Halifax Street East, Nelson, 7010, New Zealand*

2 *GNS Science, PO Box 30-368, Lower Hutt, 5040, New Zealand*

Corresponding author: Laura Biessy (laura.biessy@cawthron.org.nz)

Academic editor: Alexander Probst | **Received** 13 December 2021 | **Accepted** 1 March 2022 | **Published** 25 March 2022

Abstract

Molecular-based techniques offer considerable potential to provide new insights into the impact of anthropogenic stressors on lake ecosystems. Microbial communities are involved in many geochemical cycling processes in lakes and a greater understanding of their functions could assist in guiding more targeted remedial actions. Recent advances in metagenomics now make it possible to determine the functional potential of entire microbial communities. The present study investigated microbial communities and their functional potential in surface sediments collected from three lakes with differing trophic states and characteristics. Surface sediments were analysed for their nutrient and elemental contents and metagenomics and metabarcoding analysis undertaken. The nutrients content of the surface sediments did not show as distinct a gradient as water chemistry monitoring data, likely reflecting effects of other lake characteristics, in particular, depth. Metabarcoding and metagenomics revealed differing bacterial community composition and functional potential amongst lakes. Amongst the differentially abundant metabolic pathways, the most prominent were clusters in the energy and xenobiotics pathways. Differences in the energy metabolism paths of photosynthesis and oxidative phosphorylation were observed. These were most likely related to changes in the community composition and especially the presence of cyanobacteria in two of the three lakes. Xenobiotic pathways, such as those involving polycyclic aromatic hydrocarbons, were highest in the lakes with the greatest agricultural land-use in their catchment. These results highlight how microbial metagenomics can be used to gain insights into the causes of differences in trophic status amongst lakes.

Key Words

bacteria, cyanobacteria, New Zealand, nutrients, surface sediments, water quality

Introduction

Lakes are particularly appealing for ecological study because they essentially function as isolated island-like systems, inherently connected to their surrounding terrestrial habitats, but each lake and its catchment is usually isolated from others (Hortal et al. 2014). They are influenced by a range of physiochemical and environmental gradients and, as such, are well suited to macroecological studies (Hortal et al. 2014). Lakes cover 1.8% of the land mass on Earth, account for 0.26% of continental freshwater and are vital habitats providing essential resources for aquatic life (Scheffers and Kelletat 2016). Biological communities within lakes are affected by a variety of natural (e.g. temperature, light) and anthropogenic pressures, especially shifts in nutrient concentrations commonly associated with diffuse pollution related to land use change (Pearman et al. 2021).

Within lakes, bacteria are abundant and involved in the cycling of nutrients, playing a significant role in the breakdown of organic and inorganic compounds (Torsvik and Øvreås 2002; Zhi et al. 2015; Abia et al. 2018). These compounds can accumulate in lake sediments and can exert pressure on the microbial communities, resulting in

* These authors contributed equally.

Copyright *Laura Biessy et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

compositional changes, with possible negative impacts on lake ecosystem health (Matranga and Corsi 2012; Abia et al. 2018). Understanding how bacterial communities within lake sediments are altered by nutrient inputs is vital to predicting responses to future environmental change.

Soil and sediment contain highly diverse microbial communities (Semenov 2021). Until recently, evaluating the functions of microbial communities has been challenging and most molecular studies have focused on characterising their composition using techniques, such as metabarcoding (Caporaso et al. 2010). While metabarcoding is an extremely powerful method, it has limitations such as taxonomic biases that can be introduced by PCR primers (Wilson et al. 2019) and it does not provide any information about what metabolic functions organisms are undertaking (Semenov 2021). To address these limitations, metagenomics is now being applied to an increasing number of environmental studies (Tringe and Rubin 2005; Hugenholtz and Tyson 2008; Wilson et al. 2019). Metagenomics is a molecular technique that enables the characterisation of thousands of genes from different individual organisms within a sample (Prosser 2015; Semenov 2021). It provides a relatively unbiased overview of community structure and generates data on functional potential (Hugenholtz and Tyson 2008).

Metagenomics studies have highlighted the importance of microbial communities in nutrient cycling and biochemical degradation within aquatic ecosystems (Andreote et al. 2012; Chu et al. 2018). A metagenomics study on seawater and surface sediments from an ocean trench revealed that genes involved in chemolithoautotrophic metabolisms, such as ammonia oxidation for carbon dioxide fixation, dominated surface sediment communities (Zhang et al. 2018). Metagenomes from surface sediments of hypersaline lakes have been characterised and described and novel dominant lineages found within previously well-characterised functional groups involved in carbon, sulphur and nitrogen cycling. Key enzyme pathways were also encoded within at least four bacterial phyla never previously associated with anaerobic pathway for carbon fixation and dissimilation (Vavourakis et al. 2018).

Metagenomic studies have shown functional differences in communities across land-use, eutrophication or pollution gradients. For example, research on mangrove sediments showed reduced diazotroph abundance and nitrogen fixing capability and enhanced metabolism via increased methanogenesis and sulphate reduction in contaminated mangroves (Li et al. 2019). In addition, a high concentration of heavy metals and abundance of metal/antibiotic resistance encoding genes were found in contaminated versus pristine mangroves (Li et al. 2019). Heavily impacted marine surface sediments following an oil spill were analysed and revealed an increase in abundance of genes involved in denitrification pathways in samples that exceeded the environmental regulations for polycyclic aromatic hydrocarbons compared with those that did not (Mason et al. 2014). These results demonstrated that sediment microbiota provide an important ecosystem service for remediation of oil spills. Collectively, these metagenomics studies provide new insights into the metabolic versatility of microorganisms in aquatic sediments, their roles in carbon, nitrogen and sulphur biogeochemical cycles and how they have adapted to each unique environment. Very few metagenomic studies have been undertaken on surfaces sediments of freshwater lakes and these have only focused on the taxonomic reconstruction of ecological time series from preserved lake sediments (Garner et al. 2020) or development of a method to target microbial subpopulations (Kalyuzhnaya et al. 2008).

In this study, metagenomics was used to investigate the microbial communities and their functional potential in surface sediments collected from three lakes of differing trophic states (mesotrophic, eutrophic and supertrophic) and characteristics (e.g. size, depth, land use). The study lakes were in the Ō Tū Wharekai (Ashburton Basin, New Zealand/Aotearoa) which is an area of high cultural significance to Māori (the indigenous people of New Zealand) with high conservation values (Scott 1996; Pauling and Norton 2010). In the last two to three decades, there have been significant changes in this region including increased tourism and recreational use and intensification in farming and irrigation. The Lakes are also affected by local or regional scale climatic patterns (Bayer and Meredith 2020) and, as a result, the water quality of these Lakes has deteriorated.

The aim of this research was to describe the microbial diversity and their functional profiles in these three distinct Lakes within a small geographic area. We hypothesised that: 1) the bacterial community composition of surface sediments would differ between these Lakes; 2) KEGG IDs belonging to metabolic pathways would be differentially abundant amongst the Lakes; and 3) key differences in the functional potential of pathways involved in nutrient cycling would be observed (e.g. nitrate reduction, nitrogen fixation or denitrification, methanogenesis or sulphate reducing pathways) amongst the Lakes and these would be related to the nutrient and geochemistry of the Lakes (Abia et al. 2018; Euler et al. 2020).

Methods

Sampling sites

Three high-country (> 400 m above sea level) lakes in the Canterbury Region (New Zealand; Fig. 1), which spanned a range of trophic states and physical characteristics (Table 1), were selected for this study.

Ōtūroto/Lake Heron (Fig. 1; 43°48'79"S, 171°9'58"E) is a large, glacial formed, highland lake in the upper Rakaia River catchment (Table 1). It receives water from small streams and shingle fan seepages (Cromarty and Scott 1995) and drains to the Rakaia River. Catchment land cover is largely native vegetation or

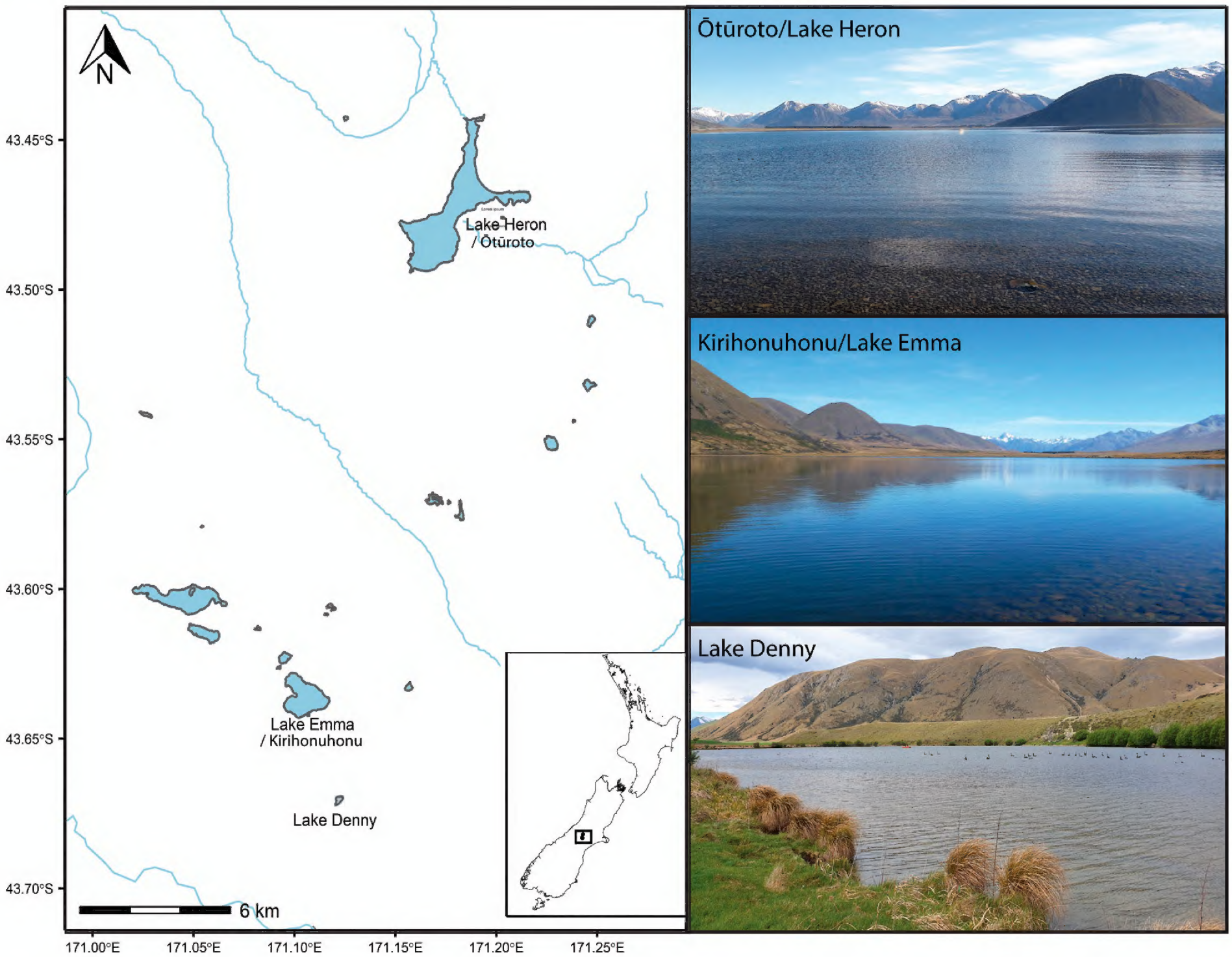

Figure 1. Locations and images of the three Lakes sampled in the Ashburton Basin (South Island, New Zealand). *Inset map* of New Zealand shows the location of the Lakes.

Table 1. Physical characteristic of each lake. Data collected from Pauling and Norton (2010) and Bayer and Meredith (2020). Res. time = Residence time. TLI = Trophic Lake Index calculated over 5 years; TP = Total phosphorus over 5 years; TN = Total nitrogen over 5 years; Chl-*a* = Chlorophyll-*a* over 5 years. Trophic Lake Index and each parameter are calculated as described in Burns et al. (1999).

	Lake type	Max. depth (m)	Altitude (m)	Area (km^2)	Res. time (y)	Seasonal stratification	Trophic level (TLI)	TP (mg/m^3)	TN (mg/m^3)	Chl-a (mg/m^3)
Lake Denny	Glacial	2.1	678	0.05	0.04	Polymictic	Supertrophic (5.52)	73	760	8.0
Kirihonuhonu/ Lake Emma	Glacial	3	640	1.67	0.11	Polymictic	Eutrophic (4.52)	22	560	7.2
Ōtūroto/Lake Heron	Glacial	37	692	6.95	0.97	Seasonal stratification	Mesotrophic (3.19)	6	145	3.9

gravel/rock, plus fringing wetlands and exotic grassland (Suppl. material 2: Table S1). Substantial areas near Ōtūroto have been converted from low to high producing grassland since the late 1990s. Ōtūroto has consistently low water column total phosphorus (TP) concentrations, but there has been an increase in chlorophyll-*a* (chl-*a*) concentrations and turbidity in the last 10 years. The Lake is currently classified as mesotrophic. Aquatic vegetation is in moderate ecological condition, associated with the widespread presence of the invasive plant *Elodea* sp. and the occasional blooms of the invasive diatom *Lindavia intermedia* ('lake snow').

Kirihonuhonu/Lake Emma (Fig. 1; 43°38'19"S, 171°6'23"E) is a medium-sized, shallow lake that drains to the south branch of the Ashburton River. In summer months, the Lake is susceptible to high nutrient levels and algal biomass. In general, nutrient levels and algal productivity have decreased in the recent years. The Lake has diverse and intact wetland fringes on the western side (Pauling and Norton 2010). Prior to the conversion of the surrounding area to conservation estate, there were reports of poor management practices, including livestock accessing lake marginal areas. Kirihonuhonu fluctuates between meso- and super-trophic, but has been mostly

eutrophic since 2016. There was a loss of macrophytes in 2009, followed by an increase in phytoplankton biomass. The invasive macrophyte *Elodea* sp. dominates and its abundance has expanded since 2007.

Lake Denny (Fig. 1; 43°40'12"S, 171°7'21"E) is a small shallow lake, draining into the Rangitata River. Lake Denny's catchments are now predominantly high and low producing exotic grasslands (Suppl. material 2: Table S1). A lack of fencing of the inflowing stream has been reported, along with instances of cattle observed in the stream, a loss of adjacent wetlands and erosion at the Lake edge. The Lake experiences regular phytoplankton blooms, very high concentrations of TN and TP, high turbidity events and high chl-*a* concentrations (Table 1). The macrophyte community was dominated by *Elodea* sp., but the Lake experienced a collapse of its macrophyte community after 2012 and was classed as 'non-vegetated' in 2017. Lake Denny has been classified as supertrophic since 2013.

Surface sediment collection for microbial DNA, nutrient and elemental characterisation

Sampling was undertaken between 4 and 9 November 2019. Prior to sampling, a side scan sonar survey was undertaken to determine the deepest part of each Lake, where subsequent sampling was undertaken. Samples for sediment geochemistry were collected by Uwitech gravity corer (90 mm diameter core). Five cores were taken and the top 2 cm of sediment removed using core cutter and combined in a 500 ml container. These were stored chilled (4 °C) and shipped to the laboratory within 48 hrs for nutrient and elemental characterisation. Ponar grab samples were taken in triplicate at each site. Using sterile spatulas, ca. 4 g of the undisturbed surface sediment layer (ca. top 1–2 mm) was placed in sterile tubes and stored in liquid nitrogen for one week before being transferred to a -80 °C freezer in the laboratory for later DNA extraction.

Nutrient and elemental samples processing

Nutrient and elemental samples were homogenised, centrifuged (3000× g, 40 min, 4 °C) and the pore water decanted. The sediments were analysed for metals using the same method as described in Pearman et al. (2020). Briefly, the sediment was dried, passed through a 2 mm sieve and analysed for the following metals using acid digestion followed by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS), based on the US Environmental Protection Agency (EPA) method 200.8 for iron (Fe), manganese (Mn), aluminium (Al), calcium (Ca), lead (Pb), copper (Cu), zinc (Zn), cadmium (Cd), phosphorus (P) and sulphur (S). Reporting limits (mg/kg) were: 12.5, 0.125, 2.5, 12.5, 0.05, 0.075, 0.05, 0.005, 10 and 250, respectively. Total nitrogen and total organic carbon (TOC) were analysed using catalytic combustion at (900 °C, O_2) and separation using a Thermal Conductivity Detector (reporting limit for both g/100 g). Oven drying of a known sediment volume was used to calculate bulk density, while oven drying followed by ashing (550 °C) and gravimetric determination was used to measure organic matter.

Metabarcoding processing

DNA for metabarcoding was extracted from 0.25 g of sediment in triplicate using the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany). The V3-V4 region of the bacterial 16S rRNA gene was amplified by polymerase chain reaction (PCR) using the primers: 341F: 5-CCT ACG GGN GGC WGC AG-3 and 805R: 5-GAC TAC HVG GGT ATC TAA TCC-3 (Herlemann et al. 2007; Klindworth et al. 2013). Conditions for the PCR reactions and subsequent library construction for sequencing on an Illumina Miseq are as described in Pearman et al. (2020). Raw reads from the sequencing machine were demultiplexed and processed using cutadapt (Martin 2011) and Amplicon Sequence Variants (ASVs) were inferred using the DADA2 (Callahan et al. 2016) using the pipeline detailed in Pearman et al. (2020). Reads were taxonomically classified using the RDP classifier (Wang et al. 2008) against the SILVA 138 database (Pruesse et al. 2007) with a minBoot threshold of 70.

Metagenomic processing

DNA was extracted from ca. 2 g of sediment using the RNeasy PowerSoil Total RNA kit (Qiagen) combined with the RNeasy PowerSoil DNA Elution kit (Qiagen) following the manufacturer's instructions. Three samples were extracted per lake. DNA was diluted to 25 µg, transferred to a DNA stable tube (GENEWIZ, Suzhou, China) and dried using a vacuum desiccator following the manufacturer's instructions before being shipped to the GENEWIZ facilities. Paired-end sequencing libraries were prepared using the VAHTS Universal DNA library kit for Illumina following the protocols of the manufacturer with amplification of the adapter-ligated DNA occurring via 8 cycles of PCR. Products were cleaned and validated using an Agilent 2100 Bioanalyzer and subsequently sequenced on an Illumina HiSeq X Ten (2ξ150 base pairs (bp)).

The raw reads from the sequencing were trimmed for adapters and removal of low quality sequences using Trimmomatic (quality thresholds: LEADING:20; TRAILING:20; SLIDINGWINDOW:40:30; MINLEN:70; (Bolger et al. 2014). Contigs were assembled separately for each sample with SPAdes (Prjibelski et al. 2020). Contigs less than 500 bp were removed from the analysis and samples were concatenated and a non-redundant catalogue of contigs was produced using cd-hit (c = 0.99). This non-redundant library of contigs was then used as the input to SqueezeMeta v.1.3.0 (Tamames and Puente-Sánchez 2019). Within the SqueezeMeta pipeline, RNAs were predicted using the software Barrnap (Seemann 2014) and 16S rRNA gene reads were classified using the same method as

for the metabarcoding. Open reading frames (ORFs) were predicted using Prodigal with the meta option selected (Hyatt et al. 2010) and the detected ORFs were functionally annotated against the Kyoto encyclopaedia of genes and genomes (KEGG; (Kanehisa and Goto 2000)) and clusters of orthologous groups of proteins (COG; (Tatusov et al. 2003)) databases using DIAMOND with the options of evalue = 0.001 and id = 30 used (Buchfink et al. 2015). Taxonomic classification of the ORFS was undertaken using DIAMOND against NCBI's nr database with the options of evalue = 0.001 and id = 40. Raw sequence reads were deposited in the National Centre for Biotechnology Information short read archive under the accession numbers PRJNA623388 and PRJNA750120 for the metagenomic and metabarcoding datasets, respectively.

Land use data

Satellite imagery available in the Land Cover Database Version 5 (LCDBv5, Landcare Research New Zealand Ltd; https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/) was used to derive seven land cover variables: (1) Native vegetation; (2) Urban; (3) Non-native vegetation; (4) Forestry; (5) High Production Grassland (HPG); (6) Low Production Grassland (LPG); and (7) Other. Suppl. material 2: Table S1 gives details of categories included in each land use variable.

Statistical analysis

Taxonomic composition of the bacterial communities was assessed by computing the relative abundance of phyla per lake with three different methods: i) 16S rRNA gene metabarcoding; ii) 16S rRNA from the metagenomic assembly and iii) taxonomic classification of the ORFs. To assess the functional aspects of the metagenome, ORFs were amalgamated to the level of KEGG ID or COG ID. The number of shared and unique identifiers (IDs) across the Lakes were calculated and visualised using the R package VennDiagram (Chen and Boutros 2011). The majority of the shared and unique annotated ORFs amongst the three Lakes using COGs annotations were assigned as 'unknown functions'; thus, the KEGG database was used in this metagenomic analysis instead of COG annotations.

Multivariate analysis of the functional data was undertaken using PERMANOVA with the R package vegan (Oksanen et al. 2019) using the factor lake for the proportional abundance of KEGG IDs, based on Bray-Curtis distance matrices. Principal Coordinate Analysis was undertaken based on the Bray-Curtis distance matrix and visualised in ggplot2 (Wickham 2016).

Differentially abundant KEGG IDs were calculated, based on the raw read numbers using the DESeq2 package in R (Love et al. 2014). KEGG IDs which had a p-value < 0.01 and had a log2 fold change of greater than 2 or less than –2 in pairwise comparisons between Lakes were considered differentially abundant. Differentially abundant KEGG IDs were depicted using ternary plots for selected pathways using the package ggtern (Hamilton and Ferry 2018).

KEGG ID abundances were summed to the pathway level for each Lake to investigate changes with environmental variables. Significant differences were assessed with Kruskal Wallis in R.

Results

Lake sediment chemistry parameters

The Lake sediments were of moderate to high density with Lake Denny being substantially higher in density than the other two Lakes. Organic matter and total organic carbon contents were relatively low and reasonably uniform across the Lakes on a per dry mass basis (Table 2). Iron and manganese contents were highest in Ōtūroto with substantially lower contents in Kirihonuhonu, while Mn:Fe molar ratios were highest in Ōtūroto and lowest in Lake Denny. The heavy metals (zinc, lead, copper and cadmium) were all highest in Lake Denny and lowest in Kirihonuhonu. In order to better allow comparison between sediments of significantly different bulk density, we also calculated organic matter and key nutrients on an areal basis (Table 3). When normalised to area, Lake Denny was substantially higher in organic content relative to Kirihonuhonu and Ōtūroto. This was also the case for nitrogen, phosphorus and sulphur contents.

Table 2. Sediment chemistry from samples of the top 0–2 cm of the lakebed sediments. Elemental results are total recoverable contents. TOC = Total organic carbon; dw = dry weight; Fe = Iron; Mn = Manganese.

Lakes	Bulk Density	Organic matter	TOC	Nitrogen	Phosphorus	Sulphur	Fe	Mn
	$kg.m^3$	g. 100 g^{-1} dw				mg.kg^{-1} dw		
Denny	284	12.9	4.90	0.57	1,040	1,200	38,100	921
Kirihonuhonu	131	18.6	5.80	0.60	730	1,800	19,100	80
Ōtūroto	141	12.8	4.70	0.64	1,900	1,100	41,600	2,370
	Aluminium	**Calcium**	**Zinc**	**Lead**	**Copper**	**Cadmium**	**Mn:Fe**	
Denny	38,500	9,070	122	34.6	61.7	0.097	0.0246	
Kirihonuhonu	17,500	18,500	49.6	11.8	18.2	0.039	0.0362	
Ōtūroto	31,000	10,500	92.7	25.9	21.5	0.074	0.0579	

Table 3. Selected sediment chemistry parameters expressed in mass per area. Elemental results are total recoverable contents.

Lakes	Areal dry mass	Organic matter	Nitrogen	Phosphorus	Sulphur
	kg.m^2		kg.m^2		
Denny	5.67	0.73	0.032	0.006	0.007
Kirihonuhonu	2.63	0.49	0.016	0.002	0.005
Ōtūroto	2.81	0.36	0.018	0.005	0.003

Metabarcoding – taxonomy analysis

In total, across the three Lakes, 275,630 reads passed the quality filtration with total prokaryotic diversity in the Lakes reaching 3,664 ASVs.

The 16S rRNA metabarcoding revealed that the prokaryotic component of the community in all three Lakes was dominated by Proteobacteria (27.6 ± 8.9% [mean ± standard deviation]; Fig. 2A) and Bacteroidota (19.8 ± 1.4%). Desulfobacterota also had a high presence across the three Lakes (12.6 ± 1.3%). Nitrospirota (mainly in the family Nitrospiraceae) had the highest relative abundance in Ōtūroto (9.9 ± 1.7%) with the lowest levels in Kirihonuhonu (3.5 ± 0.8%), while Cyanobacteria showed the opposite distribution pattern (Ōtūroto = 0.2 ± 0.05%; Kirihonuhonu = 3.4 ± 0.6%). The cyanobacteria in Kirihonuhonu and Ōtūroto mostly belonged to the family Cyanobiaceae while those in Lake Denny belonged to the family Nostocaceae with a smaller contribution from Cyanobiaceae.

Multivariate analysis showed that the taxonomic structure of the prokaryotic community in the surface sediments was significantly different, based on the 16S rRNA gene metabarcoding (*adonis*: F = 12.10; p = 0.005; Suppl. material 1: Fig. S1).

Metagenomic – taxonomy analysis

A total of 1,662 (mean = 1,230; sd = 49) 16S rRNA sequences were identified. As with the 16S rRNA metabarcoding, the Proteobacteria and Bacteroidota were the most abundant taxa account for 24.8 ± 5.3% and 17.9 ± 2.6% of the relative abundance, respectively (Fig. 2B).

A total of 9,603,406 ORFs were predicted using Prodigal from the three Lakes with an average of 4,797,666 (sd = 208,687) coding sequence ORFs per sample. Of these, 7,627,694 were assigned a taxonomy at domain level. Bacteria were the dominant component in each lake (mean 96.6%) followed by Archaea (mean 3.0%) with Eukaryota (mean 0.25%) and Viruses (mean 0.12%) accounting for only a small proportion, based on taxonomic annotation of the ORFs. As with the rRNA approaches, the bacterial component was dominated by Proteobacteria (40.5 ± 4.7%); however, Bacteroidota contributed to a substantially lower proportion of the community than with rRNA (9.0 ± 1.3%; Fig. 2C). The Archaeal component was dominated by the class Methanomicrobia (Phylum Halobacteriota; 51.3 ± 5.2% of archaea), followed by Thermoplasmata (Phylum Thermoplastmatota; 15.5 ± 6.9%) and Methanobacteria (Phylum Halobacteriota; 4.7 ± 1.4%). The eukaryotic component did not contribute significantly to the overall metagenomic community, but was predominantly classified as Annelida (Class Clitellata; 51% ± 4.7%). Lake Denny had a high contribution from Myzozoa (Class Dinophyceae; 10.1%), while Bacillariophyta reached a relative abundance of eukaryotes of 18.4% in Ōtūroto, whilst less than 2.5% in the other two Lakes.

Functional analysis

ORFs from Lake Denny had a higher average GC percentage (mean 60.6%) compared to those from Kirihonuhonu (58.3%) and Ōtūroto (58.3%). A total of 4,779,674 ORFs were annotated with KEGG (mean = 2,437,348; sd = 108,726) and were attributed to 10,867 KEGG IDs. Despite the vast majority of KEGG IDs being shared amongst the Lakes (Fig. 3A), multivariate analysis showed that there was a significant difference in the functional structure (*adonis*: F = 10.35; p = 0.003;

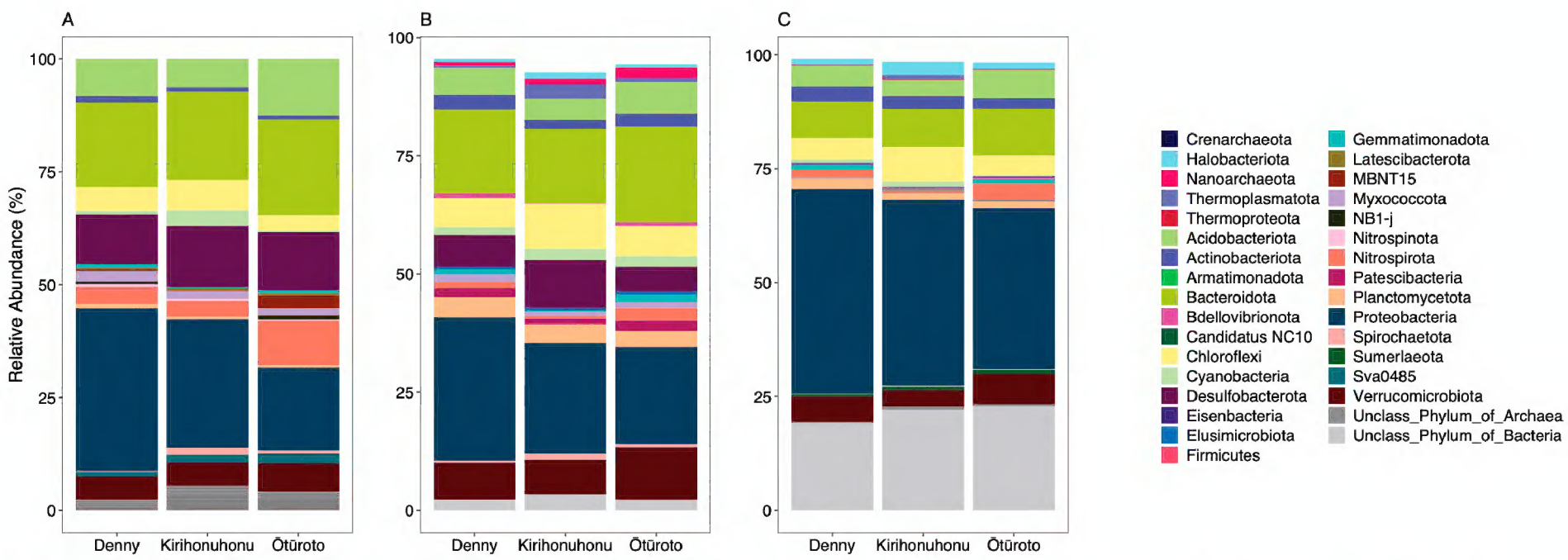

Figure 2. Taxonomic composition of the three Lakes at the phyla level for; A) 16S rRNA gene metabarcoding, B) 16S rRNA genes from the metagenomics and C) coding sequences from the metagenomics data. Only the top 20 phyla (based on mean abundance) are plotted for each method.

Fig. 3B) with Ōtūroto having the highest number of KEGG IDs not found in the other Lakes (n = 263).

In general, shared KEGG IDs in all three Lakes had similar proportions at the pathway level with carbohydrate metabolism, followed by genetic information processing and signalling and cellular processes most prominent (Suppl. material 1: Fig. S2). The unique KEGG IDs were different at the pathway level amongst Lakes. For example, genetic information processing was highest in Ōtūroto and lowest in Lake Denny with the opposite trend observed for signal transduction (Suppl. material 1: Fig. S2).

Differential abundance

In total, there were 1,326 KEGG IDs that were differentially abundant amongst the Lakes. The majority of these belonged to the KEGG pathway Metabolism with smaller numbers from the pathways Environmental Information Processing and Genetic Information Processing (Suppl. material 1: Fig. S3).

Energy metabolism had a high number of differentially abundant KEGG IDs with both Lakes Denny and Kirihonuhonu having more than Ōtūroto (Fig. 4, Suppl. material 1: Fig. S4A). KEGG IDs related to Photosynthesis and Photosynthetic antenna proteins (Fig. 5A, Suppl. material 1: Fig. S4A) had higher abundances in the shallow Lakes Denny and Kirihonuhonu. These KEGG IDs were predominantly taxonomically classified as belonging to the family Synechococcaceae (Cyanobacteria). Oxidative phosphorylation showed distinct changes amongst the Lakes. A variety of KEGG IDs related to NAD(P)H:quinone oxidoreductase were more abundant in Kirihonuhonu with abundances lowest in Ōtūroto (Fig. 5B). These KEGG IDs were taxonomically predominantly related to the family Synechococcaceae (Cyanobacteria). In contrast, Ōtūroto showed the highest abundances of the eukaryotic mitochondrial NADH:ubiquinone oxidoreductase and cytochrome c oxidase with negligible abundances in Lake Denny. The taxonomic classification of the KEGG IDs related to NADH:ubiquinone oxidoreductase and cytochrome c oxidase differed in Kirihonuhonu and Ōtūroto with the class Clitellata (phylum Annelida) being prevalent in Kirihonuhonu, while a combination of eukaryotic phototrophs (Bacillariophyta and Dinoflagellates) attributed to the taxonomy of these KEGG IDs in Ōtūroto. Other notable distributions within the energy metabolism classification of KEGG IDs were found within those that were not attributed to a metabolic pathway (unclassified pathway; Suppl. material 1: Fig. S4A) and were more prominent in Kirihonuhonu. These comprised energy-converting hydrogenase enzymes which taxonomically belonged to the archaea classes Methanobacteria and Methanomicrobia and KEGG IDs related to the biosynthesis of nitrogenase protein (Fig. 5C) attributed to Methylococcaceae and other beta- and gamma- proteobacteria.

Xenobiotics pathways were abundant in Ōtūroto and Lake Denny and were largely absent from Kirihonuhonu (Suppl. material 1: Fig. S4B). The majority of the differentially abundant KEGG IDs related to xenobiotics were for polycyclic aromatic hydrocarbons (PAHs; including styrene, toluene, xylene and naphthalene), benzoate and other pathways related to anthropogenic pressure, such as atrazine or dioxin (Suppl. material 1: Fig. S4B) and were attributed to unclassifiable taxa within the Proteobacteria.

Pathway changes associated with environmental parameters

Differences in the concentration of nutrients (nitrogen and phosphorus) per area in the Lakes were associated with changes in abundance of several metabolic pathways (Fig. 6, Suppl. material 1: Fig. S5, Suppl. material 2: Table S2).

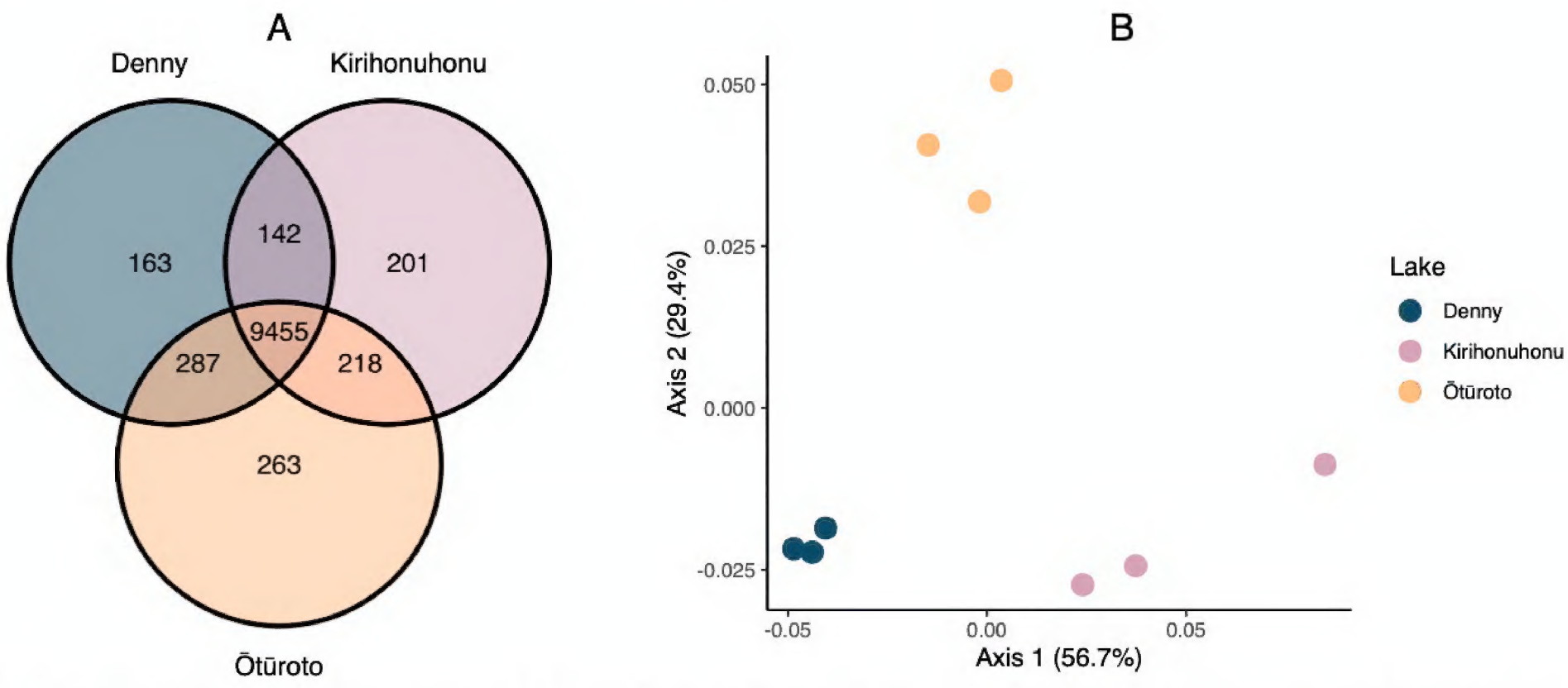

Figure 3. The shared and unique annotated Kyoto Encyclopaedia of Genes and Genomes (KEGG) IDs amongst the study three Lakes, based on Bray-Curtis dissimilarity matrix (A) and a Principal Coordinate plot of KEGG IDs, based on Bray-Curtis dissimilarity matrix.

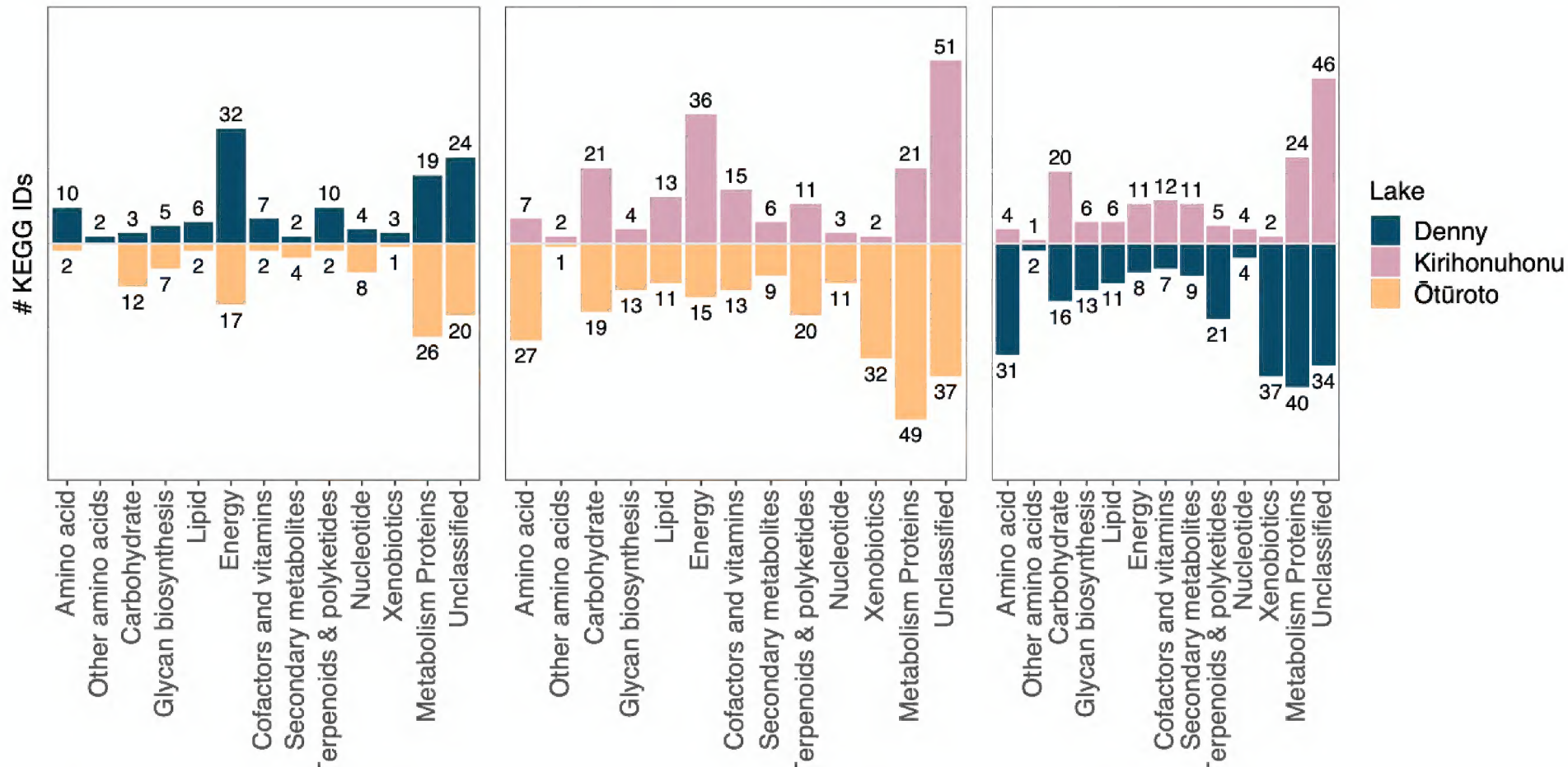

Figure 4. Number of differentially abundant Kyoto Encyclopaedia of Genes and Genomes (KEGG) IDs related to metabolism in pairwise comparisons between Lakes grouped by metabolic process.

Dissimilatory nitrate reduction to ammonia (DNRA), denitrification and nitrification pathways were all observed to have increasing abundances with an increase in the amount of nitrogen and phosphorus per m^2 (Fig. 6; Suppl. material 2: Table S2). DNRA reached a mean of 656 ± 21.2 reads per million in Lake Denny with only 433 ± 76.3 in Kirihonuhonu and were significantly different (chi squared =7.2, p=0.027). Most of the ORFs attributed to DNRA were taxonomically classified to Proteobacteria and especially the Betaproteobacteria (mean of 30.0% across the Lakes) and Deltaproteobacteria (mean of 19.3%). Although, an increase in *Nitrospira* was observed from only accounting for only 0.1% of DNRA reads in Kirihonuhonu to 2.6% in Lake Denny and 4.4% in Ōtūroto. Nitrogen fixation showed the significant (chi squared = 7.2, p = 0.027) opposite trend with higher levels (162.0 ± 14.2 reads per million) in Lake Kirihonuhonu, which has the lowest levels of nitrogen per area and lower abundances in the other two Lakes (Denny: 73.9 ± 8.9; Ōtūroto 56.3 ± 4.6). In all Lakes, ORFs related to nitrogen fixation were primarily classified as belonging to either the Deltaproteobacteria or the Gammaproteobacteria. Methanogenesis from acetate and CO_2 were both negatively associated with nitrogen and phosphorus concentrations per area with the main contributors taxonomically being the Methanomicrobia (Fig. 6; Suppl. material 2: Table S2). Assimilatory nitrate reduction was seen to increase with organic matter instead of nutrient concentrations, demonstrated with an increase in the proportion of reads attributed to cyanobacteria in the Lakes with higher organic matter (min in Ōtūroto: 0.7%; max in Kirihonuhonu at 10.8%; Fig. 7; chi squared = 6.5, p = 0.039). Assimilatory sulphate reduction was positively associated with both organic matter and sulphur per area (Fig. 7, Suppl. material 2: Table S2).

Discussion

The present study investigated the diversity and functional potential of surface sediment communities in three New Zealand lakes and related this to differences in the lake sediment chemistry and biogeochemical conditions. Water quality monitoring data indicated that Ōtūroto had the lowest trophic status and hence was less nutrient enriched than Kirihonuhonu and Lake Denny, respectively. In contrast, the dry-mass nutrient contents of the surface sediments showed a differing pattern to the water column nutrients with Ōtūroto having the highest phosphorus and nitrogen contents, likely reflecting the substantially higher iron content in the sediment of Ōtūroto, which provides strong nutrient binding capacity (Orihel et al. 2015). However, the nature of the sediments varied substantially between the Lakes, likely due to the complex interplay between catchment dynamics, lake morphometry and hydrology (Alcocer et al. 2021). For example, the sediments in Ōtūroto were low density and 'floccy' in nature compared to the higher density sediments of Lake Denny. To improve the comparison between sediments with such differing bulk densities from very different lake systems (e.g. deep/shallow), we have concentrated on area normalised parameters. As such, the supertrophic Lake Denny had the highest areal concentrations of organic matter, nutrients and sulphur.

We investigated the taxonomic composition of the surface sediments of three Lakes using a combination of metabarcoding and metagenomics. Regardless of the methodology, the phylum Proteobacteria was the dominant component of the bacterial community with Bacteroidota also contributing substantially to

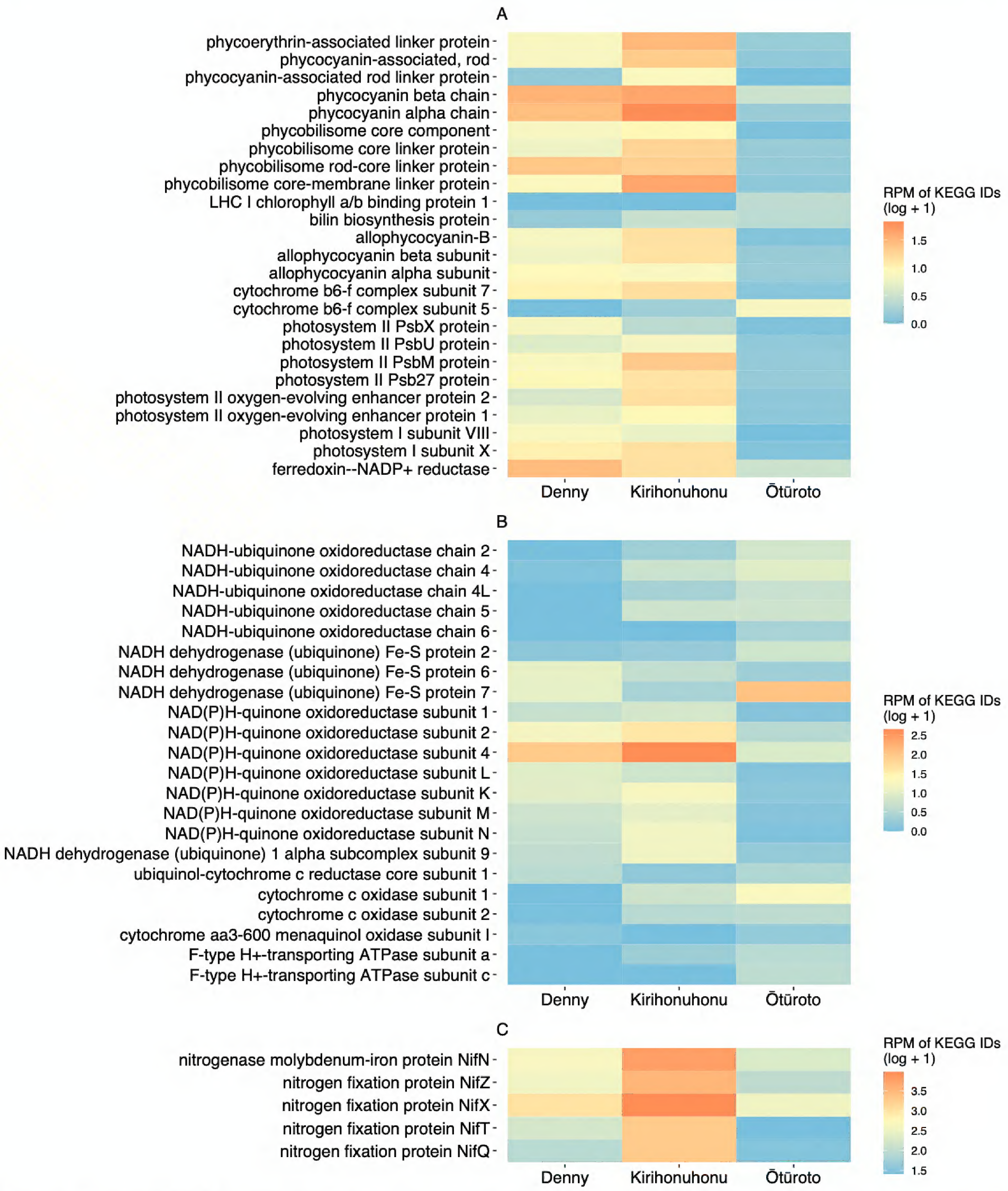

Figure 5. Heatmap of the differentially expressed genes relating to: A) the photosynthesis pathway and photosynthetic antenna proteins and B) oxidative phosphorylation. RPM = Reads per million, KEGG = Kyoto Encyclopaedia of Genes and Genomes.

the community. The dominance of these groups is in agreement with other lake sediment studies (Zhang et al. 2015), including those in New Zealand (Pearman et al. 2020). While these phyla were dominant across all Lakes, PCoA analysis of the metabarcoding results showed that there were significant ($p = 0.005$) differences amongst the Lakes in their taxonomy, confirming our first hypothesis. For example, Ōtūroto had a higher proportion of Nitrospirota and MBNT15 compared to Kirihonuhonu and Lake Denny, but a lower proportion of Cyanobacteria. Nitrospirota (previously known as Nitrospirae) are involved in nitrogen cycling, especially the conversion of nitrite to nitrate (Huang et al. 2019) and, hence, are likely indicative of oxidising conditions. Relative Mn:Fe ratios are commonly used as an indicator of redox conditions, with higher ratios indicating more

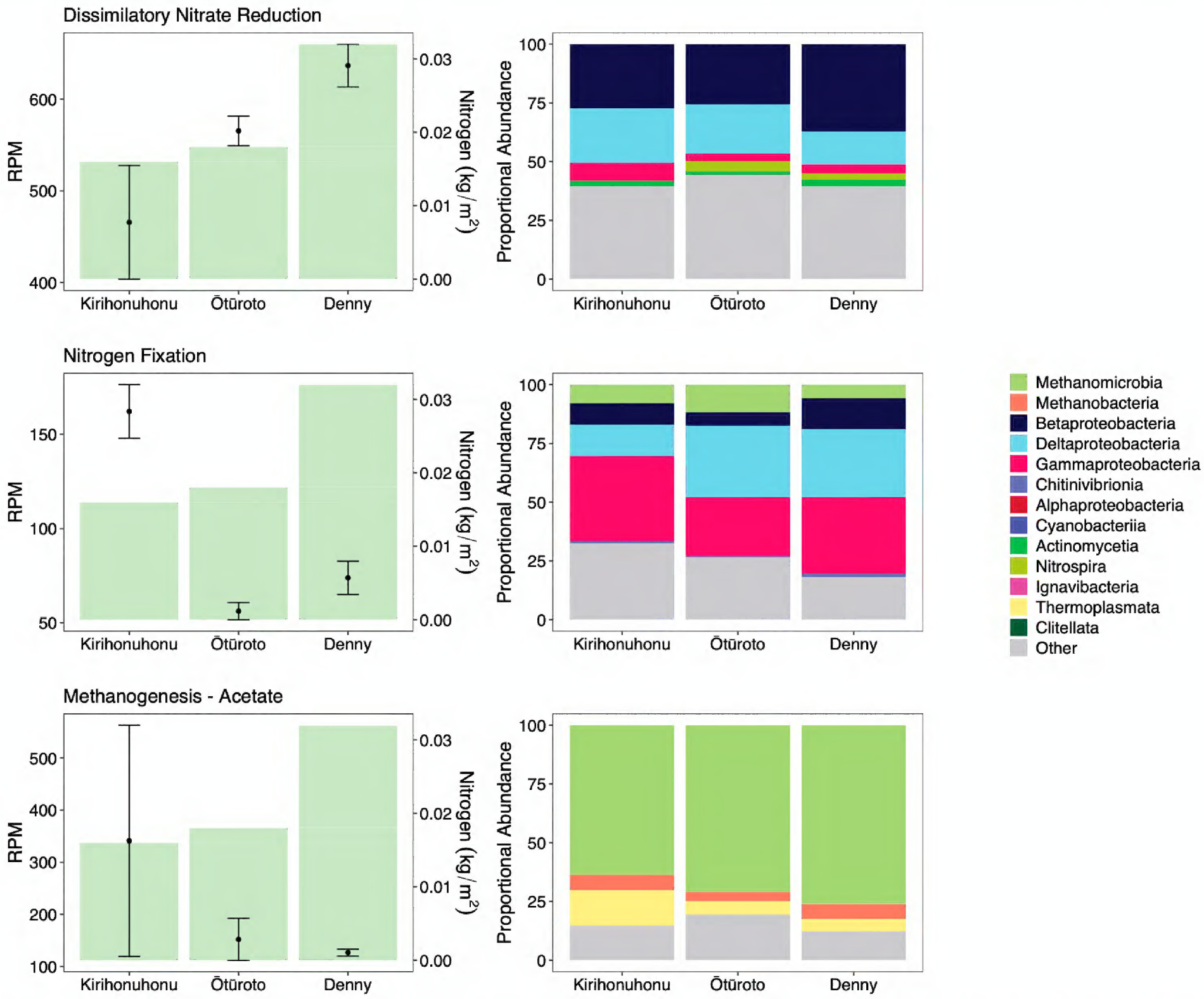

Figure 6. Abundance of metabolic pathways (reads per million; the point (mean) with standard deviation bars) per Lake and corresponding nitrogen concentrations (height of the bar plots). Nitrogen is the total nitrogen present in the sediment. Alongside is the proportional abundance of taxa for the pathway at class level. Other accounts for unclassified taxa at the class level and taxa that accounted for less than 1% of the community.

oxidising conditions (Naeher et al. 2013). Ōtūroto has the lowest relative Mn:Fe ratio of the three Lakes. The higher organic matter resulting in high sediment oxygen demand in conjunction with increased sulphur content was indicative of more reducing conditions being present in Kirihonuhonu and especially Lake Denny. This indicates that the sediment geochemistry agrees with the increased relative abundance of Nitrospirota. Taxa belonging to the MBNT15 phylum appear to be relative habitat specialists that thrive in anoxic sediments (Chen et al. 2021), although the water column of Ōtūroto is not thought to experience hypolimnetic anoxia (Bayer, pers. comm.). The higher relative abundance of cyanobacteria in Kirihonuhonu and Lake Denny was expected as they are both shallow Lakes known to experience planktonic cyanobacterial blooms (Pauling and Norton 2010). This is consistent with their trophic status and more reducing sediment geochemistry. The presence of the cyanobacteria *Nostoc* in Lake Denny also suggests that there is sufficient light reaching the benthos for it to grow. In contrast, Ōtūroto is a deeper, colder Lake where blooms are not known to occur and cyanobacterial abundances are likely to be lower.

Comparisons between the three methodologies to assess the phylogenetic composition of the Lakes showed distinct differences. Firstly, the taxonomic annotation of the ORFs had a higher relative proportion of unclassified reads at the phylum level. This is most likely due to biases in the reference databases and the lack of reference genomes present for a variety of taxa, which would limit the annotation of functional genes. This would explain why Desulfobacterota had a significant contribution to the community based on both 16S rRNA methodologies, but was absent from the community composition, based on the annotation of the ORFs. It should also be noted that the composition of a community, based on ribosomal RNA, could be biased due to the variation in copy numbers of the 16S rRNA gene amongst taxa (Kembel et al. 2012). Halobacteriota and Nanoarchaeota were observed to be more prevalent in the metagenomic-based approaches compared to the metabarcoding approach. PCR primers

Figure 7. Abundance of metabolic pathways (reads per million; the point with standard deviation bars) per Lake and corresponding organic matter concentrations (height of the bar plots). Alongside is the proportional abundance of taxa for the pathway at class level. Other accounts for unclassified taxa at the class level and taxa that accounted for less than 1% of the community.

are known to have biases in the taxa they amplify and the primer set used in this study, while it is prokaryotic specific, has been shown to have a lower coverage of archaea compared with other primers (Schlaeppi et al. 2020). Additionally, metagenomic assemblers are not optimised for the assembly of 16S rRNA genes which could impact the identification of species within the datasets (Yuan et al. 2015).

This study demonstrated that the majority of the annotated KEGG IDs (over 9,000) were shared amongst the three Lakes. This was expected as the majority of these KEGG IDs represent the core functions and metabolism of bacteria. These mainly included protein processing, carbohydrate and amino acid metabolisms, corresponding to the minimal metabolic machinery necessary for bacteria to survive (Gil et al. 2004).

We also hypothesised that the overall functional differences of the bacterial communities would vary between Lakes and that key differences in functional potential of genes involved in nutrient cycling would be observed. Multivariate analysis showed significant differences in the functional structure of the bacteria amongst the three Lakes, based on the relative abundance of KEGG IDs. While only between 1.5 and 2.4% of KEGG IDs were unique to a particular Lake, there were approximately 12% of the KEGG IDs which were differentially abundant in pair-wise comparisons of the Lakes. The majority of these differentially abundant KEGG IDs belonged to Metabolism pathways and, to a lesser extent, to Environmental Information and Genetic Information Processing pathways. Of note, were several differentially-abundant KEGG IDs clusters within the Metabolism pathways, especially within energy pathways, xenobiotics and secondary metabolites pathways. One of the main clusters of KEGG IDs abundant in Kirihonuhonu was attributed energy-converting hydrogenase and biosynthesis of nitrogenase protein. Previous studies have suggested that nitrogen fixation enzymes are negatively correlated to high nitrate concentrations (Sanseverino et al. 2021) and this was consistent with the nitrogen content of the sediments which was lowest in Kirihonuhonu. KEGG IDs, related to photosynthesis and antenna proteins, were abundant Kirihonuhonu and Lake Denny, but were almost absent from Ōtūroto. These KEGG IDs were taxonomically classified as Cyanobacteria and correlated with the higher relative abundance of the taxa in these two Lakes. Kirihonuhonu and Lake Denny are shallow Lakes with higher water column nutrient levels, conditions that promote the proliferation of cyanobacterial blooms (Wood et al. 2017), which have previously been reported in these two Lakes (Pauling and Norton 2010). There was also a difference in the relative abundance of various KEGG IDs related to oxidative phosphorylation with cyanobacterial-related KEGG IDs in Kirihonuhonu and Lake Denny. KEGG IDs, related to NADH:ubiquinone oxidoreductase, which is involved in oxidative phosphorylation, was highest in Ōtūroto and were taxonomically classified as belonging to a combination of the eukaryotic phototrophic groups, Bacillariophyta and Dinoflagellata. This may indicate that in this colder, deeper Lake where cyanobacterial blooms are not known to exist, the eukaryotic phytoplankton contribute to primary production within the Lake to a greater extent than in the more nutrient rich shallow Lakes.

Most of the KEGG IDs attributed to Xenobiotics pathways were absent or in very low relative abundance in Kirihonuhonu. Xenobiotics are synthetic chemicals from anthropogenic sources that do not or rarely exist as natural products (Rieger et al. 2002). A high proportion of the Kirihonuhonu catchment is now protected as conservation land (Bayer and Meredith 2020) and the lack of xenobiotic KEGG IDs within the surface sediment of this Lake suggests that current inputs of xenobiotics are limited. The lakes from the Ashburton Basin are popular fishing spots, but the presence of xenobiotics, such as atrazine or dioxin that are present in Lake Denny and Ōtūroto, may impact fish population. Xenobiotics, especially atrazine and certain PCBs, are well known for causing early life stage mortality in fish and for modifying physiological functions in reproductive organs, liver and neuroendocrine centres of aquatic animals (Jobling et al. 1995; Jurgella et al. 2006). Polyaromatic hydrocarbons (PAHs) are widespread environmental contaminants resulting from incomplete combustion of organic material (i.e. fossil fuels) are likely present in Ōtūroto and Lake Denny from boating or farming practices in the Lakes' surroundings. Their occurrence is an increasing concern because of their carcinogenicity and mutagenicity (Kim et al. 2013). Anthropogenic impacts are also often associated with heavy metal contamination (Duruibe et al. 2007), although other factors, such as varying geology, can also contribute. The sediments in Lake Denny, which has the most intensive catchment land use of the three Lakes in this study, showed elevated zinc, lead and copper contents relative to Kirihonuhonu and Ōtūroto. To properly quantify the relationship between the relatively recent land use change in this catchment and lake sediment contaminants, metagenomic and heavy metal analysis of a dated sediment core from the Lake would be required.

Nitrogen metabolic pathways were, in general, associated with changes in the nutrient conditions of the sediments. Dissimilatory nitrate reduction, denitrification and nitrification were all observed to increase with increasing nitrogen concentrations within the sediment. Dissimilatory nitrate reduction is in direct competition with denitrification for free nitrate (Van Den Berg et al. 2015; Broman et al. 2021) and, thus, the increase of both pathways in the more nutrient enriched lakes would be expected. Dissimilatory nitrate reduction requires anoxic conditions and, although little information is known about the oxygen conditions in Lake Denny and Kirihonuhonu, bottom waters were shown to be anoxic over a short time frame (Waters et al. 2020). However, shallow eutrophic lakes are known to experience intermittent, patchy episodes of anoxia (Waters et al. 2021) which may not be recorded in monitoring data. Hence, the occurrence of periods of anoxia in supertrophic Lake Denny would not be surprising, while the higher organic content of Kirihonuhonu sediments may explain the geochemical signature. Nitrification was also observed to increase with nutrients with increases in the proportion of Nitrospira also observed in the more nutrient enriched sites. The taxonomic classification of both steps of nitrification to Nitrospira indicates the presence of comammox (Orellana et al. 2018) in the sediment community in these more nutrient enriched Lakes, while levels were likely negligible in Kirihonuhonu where Nitrospira was less abundant. Comammox has been shown to be elevated in nutrient-rich soils (Orellana et al. 2018) and wastewater treatment plants (Yang et al. 2020) and our results support the importance of these organisms in nutrient-enriched environments. Alongside the differential abundance of biosynthesis of nitrogenase protein enzymes, the relative abundance of the nitrogen fixation pathway was highest in Kirihonuhonu which was consistent with previous work showing negative correlations of this pathway with high nitrate concentrations (Sanseverino et al. 2021). Assimilatory nitrate reduction and sulphate reduction were not correlated with nitrogen or phosphorus concentrations, but with organic matter and sulphur concentrations instead. Part of the increase in the relative abundance of assimilatory nitrate reduction could be due to the presence of cyanobacteria in Lake Denny and Kirihonuhonu which have previously been shown to undertake assimilatory nitrate reduction (Flores et al. 2005) and were shown to contribute to this pathway in these Lakes. Sulphate reduction and methanogenesis can be considered alternative degradation pathways competing for common substrates (Sinke et al. 1992). In Lake Denny, where sulphur concentrations in the sediment were highest, assimilatory sulphate reduction had higher relative abundances with methanogenesis having the lowest relative abundance of the three Lakes. This could indicate that, when sulphur and organic matter are not limited, sulphur reduction is able to compete for resources compared with methanogenesis, as has been shown in a eutrophic lake in The Netherlands (Sinke et al. 1992).

Conclusion and limitations

This small-scale metagenomic study of lake surface sediments showed the potential for such methodologies to be used to investigate functional processes occurring within lakes. The Lakes were chosen, based on monitoring data to be along a gradient of trophic state, with the microbial community related to the normalised sediment chemistry and Lake characteristics (e.g. depth). No strong differential abundance patterns of KEGG IDs were observed in key metabolic pathways (such as nitrogen reduction, sulphate reduction, methanogenesis); however, differences in photosynthetic and oxidative phosphorylation genes were recorded. These were most likely related to changes in the community and especially the presence of cyanobacteria. The overall relative abundance of metabolic pathways was shown to correlate with changes in the sediment chemistry suggesting that environmental factors are driving changes

in the functional potential of the sediment (Suppl. material 1: Fig. S6). Future studies should investigate a larger number of lakes across stronger gradients to be able to better identify changes in lake functions which could help guide lake managers on the effects of management or restoration decisions. Further metagenomics could also be used in a paleolimnological context to elucidate changes in the functional potential of bacterial communities through time and, thus, investigate how changes in the catchment (e.g. agricultural intensification) have impacted the functional potential of the lake system.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

This research was funded by the New Zealand Ministry of Business, Innovation and Employment research programme - Our lakes' health: past, present, future (C05X1707) and a Queen Elizabeth's Technician Award fund attributed to Laura Biessy. We acknowledge Te Rūnanga o Arowhenua, Te Ngāi Tūāhuriri Rūnanga, Te Rūnanga o Ngāi Tahu, Te Taumutu Rūnanga, Environment Canterbury, the Department of Conservation and landowners for their support of this project and assistance with sampling and accessing the sites. We thank Jonathan Puddick (Cawthron Institute), Chris Moy (University of Otago), Julia Short (University of Adelaide) and Riki Ellison (Waka Taurua Ltd) for field assistance and Chris Weisener and Nick Falk (University of Windsor) for advice during project planning. The Department of Conversation is acknowledged for assistance with permitting. We would like to thank the editor and two anonymous reviewers for their input in improving the manuscript.

References

Abia ALK, Alisoltani A, Keshri J, Ubomba-Jaswa E (2018) Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. The Science of the Total Environment 616: 326–334. https://doi.org/10.1016/j.scitotenv.2017.10.322

Alcocer J, Prado B, Mora L, Oseguera LA, Caballero M (2021) Sediment characteristics of tropical, karst lakes and their relationship with watershed topography, lake morphometry, and human activities. Journal of Paleolimnology 66(3): 333–353. https://doi.org/10.1007/s10933-021-00210-z

Andreote FD, Jiménez DJ, Chaves D, Dias ACF, Luvizotto DM, Dini-Andreote F, Fasanella CC, Lopez MV, Baena S, Taketani RG, de Melo IS (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. PLoS ONE 7(6): e38600. https://doi.org/10.1371/journal.pone.0038600

Bayer T, Meredith A (2020) Canterbury high-country lakes monitoring programme - state and trends, 2005–2019. Environment Canterbury Regional Council, 224 pp.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Broman E, Zilius M, Samuiloviene A, Vybernaite-Lubiene I, Politi T, Klawonn I, Voss M, Nascimento FJ, Bonaglia S (2021) Active DNRA and denitrification in oxic hypereutrophic waters. Water Research 194: e116954. https://doi.org/10.1016/j.watres.2021.116954

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12(1): 59–60. https://doi.org/10.1038/nmeth.3176

Burns NM, Rutherford JC, Clayton JS (1999) A monitoring and classification system for New Zealand lakes and reservoirs. Lake and Reservoir Management 15(4): 255–271. https://doi.org/10.1080/07438149909354122

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5): 335–336. https://doi.org/10.1038/nmeth.f.303

Chen H, Boutros PC (2011) VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12(1): 1–7. https://doi.org/10.1186/1471-2105-12-35

Chen Y-J, Leung PM, Wood JL, Bay SK, Hugenholtz P, Kessler AJ, Shelley G, Waite DW, Franks AE, Cook PL, Greening C (2021) Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. The ISME Journal 15(10): 1–19. https://doi.org/10.1038/s41396-021-00988-w

Chu BT, Petrovich ML, Chaudhary A, Wright D, Murphy B, Wells G, Poretsky R (2018) Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. Applied and Environmental Microbiology 84(5): e02168-17. https://doi.org/10.1128/AEM.02168-17

Cromarty P, Scott DA (1995) A directory of wetlands in New Zealand. Department of Conservation, Wellington, New Zealand, 394 pp.

Duruibe JO, Ogwuegbu M, Egwurugwu J (2007) Heavy metal pollution and human biotoxic effects. International Journal of Physical Sciences 2: 112–118.

Euler S, Jeffrey LC, Maher DT, Mackenzie D, Tait DR (2020) Shifts in methanogenic archaea communities and methane dynamics along a subtropical estuarine land use gradient. PLoS ONE 15(11): e0242339. https://doi.org/10.1371/journal.pone.0242339

Flores E, Frías JE, Rubio LM, Herrero A (2005) Photosynthetic nitrate assimilation in cyanobacteria. Photosynthesis Research 83(2): 117–133. https://doi.org/10.1007/s11120-004-5830-9

Garner RE, Gregory-Eaves I, Walsh DA (2020) Sediment metagenomes as time capsules of lake microbiomes. MSphere 5(6): 512–520. https://doi.org/10.1128/mSphere.00512-20

Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. Microbiology and Molecular Biology Reviews 68(3): 518–537. https://doi.org/10.1128/MMBR.68.3.518-537.2004

Hamilton NE, Ferry M (2018) ggtern: Ternary diagrams using ggplot2. Journal of Statistical Software 87(Code Snippet 3): 1–17. https://doi.org/10.18637/jss.v087.c03

Herlemann DPR, Geissinger O, Brune A (2007) The termite group I phylum is highly diverse and widespread in the environment. Applied and Environmental Microbiology 73(20): 6682–6685. https://doi.org/10.1128/AEM.00712-07

Hortal J, Nabout JC, Calatayud J, Carneiro FM, Padial A, Santos A, Siqueira T, Bokma F, Mauricio Bini L, Ventura M (2014) Perspectives on the use of lakes and ponds as model systems for macroecological research. Journal of Limnology 73(s1): 46–60. https://doi.org/10.4081/jlimnol.2014.887

Huang W, Chen X, Wang K, Chen J, Zheng B, Jiang X (2019) Comparison among the microbial communities in the lake, lake wetland, and estuary sediments of a plain river network. MicrobiologyOpen 8(2): e00644. https://doi.org/10.1002/mbo3.644

Hugenholtz P, Tyson GW (2008) Metagenomics. Nature 455(7212): 481–483. https://doi.org/10.1038/455481a

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11(1): 1–11. https://doi.org/10.1186/1471-2105-11-119

Jobling S, Reynolds T, White R, Parker MG, Sumpter JP (1995) A variety of environmentally persistent chemicals, including some phthalate plasticizers, are weakly estrogenic. Environmental Health Perspectives 103(6): 582–587. https://doi.org/10.1289/ehp.95103582

Jurgella GF, Marwah A, Malison JA, Peterson R, Barry TP (2006) Effects of xenobiotics and steroids on renal and hepatic estrogen metabolism in lake trout. General and Comparative Endocrinology 148(2): 273–281. https://doi.org/10.1016/j.ygcen.2006.03.011

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suciu D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L (2008) High-resolution metagenomics targets specific functional types in complex microbial communities. Nature Biotechnology 26(9): 1029–1034. https://doi.org/10.1038/nbt.1488

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28(1): 27–30. https://doi.org/10.1093/nar/28.1.27

Kembel SW, Wu M, Eisen JA, Green JL (2012) Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Computational Biology 8(10): e1002743. https://doi.org/10.1371/journal.pcbi.1002743

Kim K-H, Jahan SA, Kabir E, Brown RJ (2013) A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. Environment International 60: 71–80. https://doi.org/10.1016/j.envint.2013.07.019

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research 41(1): e1. https://doi.org/10.1093/nar/gks808

Li Y, Zheng L, Zhang Y, Liu H, Jing H (2019) Comparative metagenomics study reveals pollution induced changes of microbial genes in mangrove sediments. Scientific Reports 9(1): 1–11. https://doi.org/10.1038/s41598-019-42260-4

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12): e550. https://doi.org/10.1186/s13059-014-0550-8

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet 17(1): 10–12. https://doi.org/10.14806/ej.17.1.200

Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Bælum J, Kimbrel J, Bouskill NJ, Prestat E, Borglin S, Joyner DC, Fortney JL, Jurelevicius D, Stringfellow WT, Alvarez-Cohen L, Hazen TC, Knight R, Gilbert JA, Jansson JK (2014) Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. The ISME Journal 8(7): 1464–1475. https://doi.org/10.1038/ismej.2013.254

Matranga V, Corsi I (2012) Toxic effects of engineered nanoparticles in the marine environment: Model organisms and molecular approaches. Marine Environmental Research 76: 32–40. https://doi.org/10.1016/j.marenvres.2012.01.006

Naeher S, Gilli A, North RP, Hamann Y, Schubert CJ (2013) Tracing bottom water oxygenation with sedimentary Mn/Fe ratios in Lake Zurich, Switzerland. Chemical Geology 352: 125–133. https://doi.org/10.1016/j.chemgeo.2013.06.006

Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara R, Simpson G, Solymos P (2019) Vegan: Community ecology package, 298 pp.

Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT (2018) Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Applied and Environmental Microbiology 84(2): e01646-17. https://doi.org/10.1128/AEM.01646-17

Orihel DM, Schindler DW, Ballard NC, Graham MD, O'Connell DW, Wilson LR, Vinebrooke RD (2015) The "nutrient pump:" Iron-poor sediments fuel low nitrogen-to-phosphorus ratios and cyanobacterial blooms in polymictic lakes. Limnology and Oceanography 60(3): 856–871. https://doi.org/10.1002/lno.10076

Pauling C, Norton T (2010) Ō Tū Wharekai ora tonu. Cultural health assessment of Ō Tū Wharekai /The Ashburton Lakes. Te Rūnanga O Arowhenua, 84 pp.

Pearman JK, Biessy L, Thomson-Laing G, Waters S, Vandergoes MJ, Howarth JD, Rees A, Moy C, Pochon X, Wood SA (2020) Local factors drive bacterial and microeukaryotic community composition in lake surface sediment collected across an altitudinal gradient. FEMS Microbiology Ecology 96(6): 70–83. https://doi.org/10.1093/femsec/fiaa070

Pearman JK, Thomson-Laing G, Howarth JD, Vandergoes MJ, Thompson L, Rees A, Wood SA (2021) Investigating variability in microbial community composition in replicate environmental DNA samples down lake sediment cores. PLoS ONE 16(5): e0250783. https://doi.org/10.1371/journal.pone.0250783

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. Current Protocols in Bioinformatics 70(1): e102. https://doi.org/10.1002/cpbi.102

Prosser JI (2015) Dispersing misconceptions and identifying opportunities for the use of'omics' in soil microbial ecology. Nature Reviews. Microbiology 13(7): 439–446. https://doi.org/10.1038/nrmicro3468

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35(21): 7188–7196. https://doi.org/10.1093/nar/gkm864

Rieger P-G, Meier H-M, Gerle M, Vogt U, Groth T, Knackmuss H-J (2002) Xenobiotics in the environment: Present and future strategies to obviate the problem of biological persistence. Journal of Biotechnology 94(1): 101–123. https://doi.org/10.1016/S0168-1656(01)00422-9

Sanseverino I, Pretto P, António DC, Lahm A, Facca C, Loos R, Skejo H, Beghi A, Pandolfi F, Genoni P, Lettieri T (2021) Metagenomics analysis to investigate the microbial communities and their functional profile during cyanobacterial blooms in Lake Varese. Microbial Ecology: 1–19. https://doi.org/10.1007/s00248-021-01914-5

Scheffers AM, Kelletat DH (2016) Lakes of the world with google earth: understanding our environment. Springer International Publishing, 303 pp. https://doi.org/10.1007/978-3-319-29617-3

Schlaeppi K, Ronchi F, Leib SL, Erb M, Ramette A (2020) Evaluation of primer pairs for microbiome profiling from soils to humans within the One Health framework. Molecular Ecology Resources 20(6): 1558–1571. https://doi.org/10.1111/1755-0998.13215

Scott DA (1996) A Directory of Wetlands in New Zealand. 1st Edn. Department of Conservation, 43 pp.

Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. Bioinformatics (Oxford, England) 30(14): 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Semenov M (2021) Metabarcoding and metagenomics in soil ecology research: Achievements, challenges, and prospects. Biology Bulletin Reviews 11(1): 40–53. https://doi.org/10.1134/S2079086421010084

Sinke AJ, Cornelese AA, Cappenberg TE, Zehnder AJ (1992) Seasonal variation in sulfate reduction and methanogenesis in peaty sediments of eutrophic Lake Loosdrecht, The Netherlands. Biogeochemistry 16(1): 43–61. https://doi.org/10.1007/BF02402262

Tamames J, Puente-Sánchez F (2019) SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. Frontiers in Microbiology 9: e3349. https://doi.org/10.3389/fmicb.2018.03349

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: An updated version includes eukaryotes. BMC Bioinformatics 4(1): 1–14. https://doi.org/10.1186/1471-2105-4-41

Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: From genes to ecosystems. Current Opinion in Microbiology 5(3): 240–245. https://doi.org/10.1016/S1369-5274(02)00324-7

Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. Nature Reviews. Genetics 6(11): 805–814. https://doi.org/10.1038/nrg1709

Van Den Berg EM, Van Dongen U, Abbas B, Van Loosdrecht MC (2015) Enrichment of DNRA bacteria in a continuous culture. The ISME Journal 9(10): 2153–2161. https://doi.org/10.1038/ismej.2015.26

Vavourakis CD, Andrei A-S, Mehrshad M, Ghai R, Sorokin DY, Muyzer G (2018) A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome 6(1): 1–18. https://doi.org/10.1186/s40168-018-0548-7

Wang X-J, Yu R-C, Luo X, Zhou M-J, Lin X-T (2008) Toxin-screening and identification of bacteria isolated from highly toxic marine gastropod *Nassarius semiplicatus*. Toxicon 52(1): 55–61. https://doi.org/10.1016/j.toxicon.2008.04.170

Waters S, Webster-Brown JG, Hawes I (2020) The release of legacy phosphorus from deforestation-derived sediments in shallow, coastal Lake Forsyth/Te Roto o Wairewa. New Zealand Journal of Marine and Freshwater Research 55(3): 446–465. https://doi.org/10.1080/00288330.2020.1804408

Waters S, Verburg P, Schallenberg M, Kelly D (2021) Sedimentary phosphorus in contrasting, shallow New Zealand lakes and its effect on water quality. New Zealand Journal of Marine and Freshwater Research 55(4): 592–611. https://doi.org/10.1080/00288330.2020.1848884

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer International Publishing, 260 pp. https://doi.org/10.1007/978-3-319-24277-4_9

Wilson J-J, Brandon-Mong G-J, Gan H-M, Sing K-W (2019) High-throughput terrestrial biodiversity assessments: Mitochondrial metabarcoding, metagenomics or metatranscriptomics? Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis 30(1): 60–67. https://doi.org/10.1080/24701394.2018.1455189

Wood SA, Maier MY, Puddick J, Pochon X, Zaiko A, Dietrich DR, Hamilton DP (2017) Trophic state and geographic gradients influence planktonic cyanobacterial diversity and distribution in New Zealand lakes. FEMS Microbiology Ecology 93(2): 1–13. https://doi.org/10.1093/femsec/fiw234

Yang Y, Daims H, Liu Y, Herbold CW, Pjevac P, Lin J-G, Li M, Gu J-D (2020) Activity and metabolic versatility of complete ammonia oxidizers in full-scale wastewater treatment systems. ASM Journals : mBio 11(2): e03175-19. https://doi.org/10.1128/mBio.03175-19

Yuan C, Lei J, Cole J, Sun Y (2015) Reconstructing 16S rRNA genes in metagenomic data. Bioinformatics (Oxford, England) 31(12): 35–43. https://doi.org/10.1093/bioinformatics/btv231

Zhang J, Yang Y, Zhao L, Li Y, Xie S, Liu Y (2015) Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. Applied Microbiology and Biotechnology 99(7): 3291–3302. https://doi.org/10.1007/s00253-014-6262-x

Zhang X, Xu W, Liu Y, Cai M, Luo Z, Li M (2018) Metagenomics reveals microbial diversity and metabolic potentials of seawater and surface sediment from a hadal biosphere at the Yap Trench. Frontiers in Microbiology 9: e2402. https://doi.org/10.3389/fmicb.2018.02402

Zhi E, Song Y, Duan L, Yu H, Peng J (2015) Spatial distribution and diversity of microbial community in large-scale constructed wetland of the Liao River Conservation Area. Environmental Earth Sciences 73(9): 5085–5094. https://doi.org/10.1007/s12665-015-4021-7

Supplementary material 1

Figures S1–S6

Author: Laura Biessy, John K. Pearman, Sean Waters, Marcus J. Vandergoes, Susanna A. Wood

Data type: images

Explanation note: **Fig. S1.** Principal Coordinate plot of the 16S rRNA gene metabarcoding results based on Bray-Curtis distances. **Fig. S2.** Heatmap of the differentially abundant genes. **Fig. S3.** Number of differentially abundant Kyoto Encyclopaedia of Genes and Genomes (KEGG). **Fig. S4.** Ternary plots illustrating the Kyoto Encyclopaedia of Genes and Genomes (KEGG). **Fig. S5.** Abundance of metabolic pathways. **Fig. S6.** Review figure showing the responses of metabolic pathways and nutrient concentrations in the three lakes.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.79265.suppl1

Supplementary material 2
Tables S1, S2

Author: Laura Biessy, John K. Pearman, Sean Waters, Marcus J. Vandergoes, Susanna A. Wood

Data type: docx

Explanation note: **Table S1.** Land cover variables used and the categories included, and percentage of the catchment in each land use cover for the study lakes. **Table S2.** The reads per million (RPM) with standard deviation (S.D) for different metabolic pathways per lake association with different environmental variables.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.79265.suppl2